DOCUMENT RESUME

ED 148 741 SP 011 646

AUTHOR Medley, Donald M.
TITLE Evaluation Designs for Innovative Teacher Education Programs. Comments.
PUB DATE 4 Apr 72
NOTE 11p.; Presented before the American Educational Research Association (Chicago, Illinois, April 4, 1972)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Educational Research; *Effective Teaching; *Evaluation Needs; Program Evaluation; *Research Design; Research Needs; *Student Improvement
IDENTIFIERS *Critical Analysis

ABSTRACT

The author criticizes a number of research designs in which he finds two recurring defects: (1) the differentiation between research and evaluation, with the belief that the two concepts are incompatible, and that if one objective is chosen as the field of action, the other must be avoided; and (2) the nature of evaluation itself. This second defect is evidenced in researcher avoidance of testing student outcomes to validate research findings on teacher improvement. It is suggested that researchers and evaluators are under a strong obligation to collect pupil learning data as a part of evaluation projects. (MJB)

**********
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
**********

For your information, Al

DMM PB0920

# EVALUATION DESIGNS FOR INNOVATIVE TEACHER EDUCATION PROGRAMS

file

Critique, SIG/Teacher Preparation Program

American Educational Research Association

Chicago, April 4, 1972

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Donald M. Medley
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND THE ERIC SYSTEM CONTRACTORS."

COMMENTS

Donald M. Medley

Performance
Educati
A

ED148741

I am going to take a somewhat negative stance today and criticize the three designs we have heard about for two defects which they all share to a greater or less degree. In doing so I shall ignore many strong points in the designs, so I hope you will not interpret my comments as indicating a low overall evaluation of any of them.

The defects which concern me are almost universal in current efforts to evaluate teacher education programs of all types, growing, as they do, out of misconceptions that are widely held and widely taught. One misconception has to do with the difference between research and evaluation; and the other has to do with the nature of evaluation itself.

A particularly clear statement of the position often taken on the difference between research in teacher education and the evaluation of teacher education programs has been supplied by Dr. Greenberg and his colleagues in their report on plans for evaluating the extensive Maryland teacher education center program. They define research as an activity "concerned with the production of new knowledge." Nothing wrong with that, certainly. And they define evaluation as an activity which "has as its primary focus

SP011 646

the determination of either absolute or comparative worth." Nothing wrong with that either.

What I object to is the general notion that these two activities are somehow incompatible--that you must decide which you are going to do, and having decided, eschew the other. According to our definitions, the choice you have is either to develop new knowledge or to determine the value of something. You can't do both.

I must hasten to admit that Greenberg tempers this somewhat in his report by saying that evaluation may contribute to the advancement of knowledge; but he goes on to warn us that in merging the two we run the risk of doing neither task adequately.

This puzzled me the first time I heard it. Except in the rare case when all of the knowledge necessary for evaluating a program exists and is at hand, the development of new knowledge would seem a necessary part of evaluation, since it would be needed to provide the very factual base on which the evaluation is based. Any attempt to evaluate without developing some new knowledge--in other words, without also doing some research--must fail. To do an adequate piece of evaluation one would, it seems to me, have to do an adequate piece of research, too.

The suggestion made in the same paragraph that combining the two efforts is likely to interfere with the evaluation process seems to relate to a difference in generalizability of findings of the two kinds of activity. When we develop knowledge to assist decision-making in a particular setting, that knowledge may not apply very well to other settings. That is true, and the

generalizability of such findings may therefore be limited, at least in the strict sense of statistical inference.

But does anyone here think that the situation is very different in regard to findings of the typical piece of research in teacher education? There is, in fact, a strong tradition of overgeneralization of research findings. There are teacher education programs which contain large elements based on the application of knowledge developed by research studies done in pigeon laboratories and T-group sessions to classroom behavior. Having swallowed elephants like this, should we suddenly strain at a gnat? Should we say that we cannot apply what was found in a Maryland teacher education center to a field-centered program in Oregon because the data were collected primarily for local use? Believe me, mixing research and evaluation does not present any real danger of lowering the quality of the research. And since evaluation must be based on research, only bad research can lower the quality of evaluation.

I have heard this idea that research and evaluation don't mix before, and am curious as to how it has come to be so widely believed. It seems to be related to the even stranger myth that somebody somewhere is doing some kind of pure research in teacher education, and that one of these days we are going to get all our answers from those boys. I think this myth survives mainly because it is useful. It gives us an excuse for doing what we are doing in teacher education, and for not doing better. While we are waiting for the researchers to catch up with us and tell us what competencies teachers need, we can do whatever we

think best with no responsibility for showing that what we are doing is really improving education. This is not really a very comfortable position, but its the best we can do.

My second objection to the way we evaluate teacher education programs nowadays has to do with another of these escape routes, one based on the fact that so long as we are trying to do the impossible, nobody can blame us for failing! What I am talking about here is the general agreement that the ultimate criterion against which any teacher education program must be evaluated is how much pupils learn from teachers trained in the program. It is hard to disagree with this.

But the trouble is that it is practically impossible to carry this idea out in practice. To demonstrate that pupils would learn more from a teacher trained in a particular way than they would have learned from him if he had not been trained that way is far beyond our present capability, and we know it. As a matter of fact, there is very little evidence--and what there is is highly questionable--that what a teacher does has anything to do with how much a pupil learns, let alone that part of what the teacher does that is attributable to his training has any impact. Now I am not saying that teachers do not have an impact, and an important one; but I am saying that there is not much research to prove it. There is little evidence for the simple reason that such evidence is almost impossible to come by.

If we set out to evaluate a program--the Weber State one, let us say--on the basis of how much more effective teachers

trained in it are in the classroom than they would have been if they had not been trained at Weber State, we are undertaking to develop some very new knowledge indeed! But do we do that? Oh no. Not us. We know better than that. We say that in a few years from now that is what we will do. Right now we will operate at a provisional level somewhere short of what everybody agrees is really the only right way to do it. Once again we avoid the final crucial test. This time we do not leave it to the researchers (whoever they are), but we just postpone it.

These, then, somewhat oversimplified and overstated, are my two objections to the generally accepted model for evaluating teacher education. The research that will tell us what we should be teaching our students is supposed to be done by somebody else--some researchers who really do not exist. The evaluation of our teacher education programs against our ultimate goal is postponed indefinitely while we go on with business as usual. As long as we go on this way, and feel comfortable in doing so, we will never really come to grips with our basic problem of how to improve teaching, and never move toward its solution.

I have the feeling that after stating the problem in such strong terms, anything I suggest in the way of a solution is going to look pretty anti-climactic. Well, so be it. Be prepared for a let-down, because what I have to propose is not very dramatic or revolutionary, and is not going to change things overnight. But it might get things moving.

Let me present the basic idea by using the evaluation of an institute for mathematics teachers described here today by Hansen and Pikaart as an example. They have used the standard evaluation design in its best form, and with about as favorable results as you could reasonably hope to get.

When teachers who took part in the institute were compared with ones who did not on mastery of the mathematical concepts taught, they were found to know more on the average. This, as the authors point out, was an evaluation on what Turner would call level 6. This was entirely appropriate, since the immediate objectives of the institute were defined on that level. Let me put it this way:

The institute was designed to raise the level of knowledge of mathematics possessed by elementary school teachers, under the hypothesis that pupils will learn more mathematics from teachers who know more about mathematics.

The institute was successful in achieving its objective.

In the prescribed fashion, the investigators then proceeded to evaluate the program in terms of the ultimate criterion by comparing mean gains of pupils taught by the two groups of teachers. Not surprisingly, no significant difference was found.

Does this mean that we should evaluate the program as ineffective? Should the decision have been made to discontinue the institute because it did not increase the effectiveness of teachers trained in it? Theoretically, yes. But practically we know that if we abandoned every institute and program that could

not be shown to increase teacher effectiveness, we would all be out of business. So instead we say that, although the null hypothesis was not disproven, it was not proven either; and decide that the program is worth continuing until we can think of something better to try.

This is the generally recommended way of proceeding, and this is what generally happens. We really did not learn as much from the project as we would have liked--we did not really develop any new knowledge on which we could base an evaluation of the program that could be defended under the assumption that a good program must increase pupil learning. In other words, we did not find out whether it was a good or a bad program.

The alternative I want to suggest involves two changes--one in interpretation, and one in procedure.

Let me repeat the statement of the problem as I formulated it just now. I said that the institute was designed to raise the level of knowledge of elementary school teachers, under the hypothesis that pupils will learn more mathematics from teachers who know more of the concepts taught in the institute.

I suggest, first, that we interpret the finding that the institute did achieve its objective of raising the average level of mathematical knowledge of the teachers who participated in it as meaning that the institute was successful. I suggest that this finding be shared with anyone else who wants to achieve the same objectives, whether because he accepts the hypothesis on which the institute was based as true or not. If you want to increase

teachers' knowledge of mathematics, our evaluation says this is what we tried here and it worked. Whether someone else wants to apply what we learned in a different situation is his business. As research results, these may not be the most compelling or startling in the world, or those with the widest implications. But they do constitute an increment to knowledge about the education of elementary school teachers. That is what I mean by a change in interpretation.

The change in procedure that I want to propose is that we use the data already collected to test the hypothesis that mathematical knowledge is related to pupil learning. I know that some would say that by suggesting that we do some research, I expose us to the risk of spoiling our evaluation. But I don't see how. I see us developing new knowledge, true. But that new knowledge can be highly relevant to evaluation, and even improve it. Because if we can show such a relationship, then within the limitations of these data we will have evidence that the institute program is improving teacher effectiveness. Mind now that we will not have shown that our experimental group is more effective than our control group; but we will have shown that the more an individual teacher has of what the institute is designed to produce, the more his pupils are likely to learn; and we will also have shown that participating in the institute gives teachers more of this knowledge.

The prospects for establishing such a relationship are considerably better than those for finding a difference between

the means of the two groups. The two-group analysis uses institute membership as a measure of knowledge of mathematical concepts, not test scores; and it treats all teachers who took part in it as if they all had equal knowledge (except for the stratification). It also treats all control teachers as though they had equal knowledge. I am sure that there was in fact a considerable range of knowledge within each group, as well as a substantial amount of overlap between groups. This means that teachers who were equal in knowledge but in different groups were counted as different in knowledge, and that teachers who differed considerably in knowledge but were in the same group were treated as equal in knowledge. And the same thing applies to the measure of pupil learning: all pupils whose teachers attended the institute were regarded as identical in amount learned; and all whose teachers did not attend were likewise treated as identical.

When we correlate teacher knowledge and mean pupil gain, our measure of knowledge ignores group membership and treats differences as differences wherever they occur, and our measure of pupil learning does likewise. The sensitivity or power of this design is much greater than the two-groups one. It may still not be sensitive enough to detect a relationship between a level 6 criterion and pupil gains, but at least there is a chance.

I told you there was nothing very startling about this model. All I am saying is that a program be evaluated on its effectiveness in producing the change in teachers it is supposed to produce, and that the relationship of the changes to effects on pupils be tested independently of this evaluation.

In this example, my model can be applied without a need for collecting any data not called for by the conventional model. This should be true for any evaluation in which pupil learning is measured. If a training activity is designed to develop a competency in teachers that has not yet been shown to relate to measured teacher effectiveness, the evaluator is under a strong obligation to collect pupil learning data as part of his evaluation project. If this obligation is taken seriously in the future, we ought to begin very soon to accumulate knowledge about which competencies our programs should be developing in order to turn out more effective teachers.

As long as we leave this problem to the researchers, it will remain unsolved; and it is certainly the central problem in teacher education. I have tried to suggest to you that there are no researchers out there who can solve this problem for us. We are going to have to solve it ourselves. We must do the research ourselves, or it will not get done. If I can stretch the familiar Pogo-ism a bit, I might suggest as a motto for program evaluators:

We have met the researcher, and he is us.